2016年6月17日　統計数理研究所　オープンハウス

# 高頻度金融データのモデリングと統計解析

荻原　哲平　　数理・推論研究系　助教

## 1. 高頻度金融データのモデリング

株式市場の株価データからその変動の大きさや異なる株式間での株価の連動性などの特性を統計的に解析することは、株式資産のリスクを管理する上で重要な課題と言える. 従来の金融資産のリスク管理においては, 一日の終わりの株価を用いるなどの日単位の株価データを用いた分析が主流だったが, 近年は一日内の全ての株式取引のデータに関する情報がデータベースで管理されるようになり, それらを用いた研究が活発になっている. それらは秒・ミリ秒単位の高頻度に観測されたデータであり, 日単位のデータに比べて非常に多くの情報を含んでいる. このような高頻度金融データのモデリングとして,「拡散過程による株価モデリング」と「点過程を用いた株式板情報のモデリング」を考える.

## 2. 拡散過程による株価モデリング

### 【高頻度金融データ特有の問題】

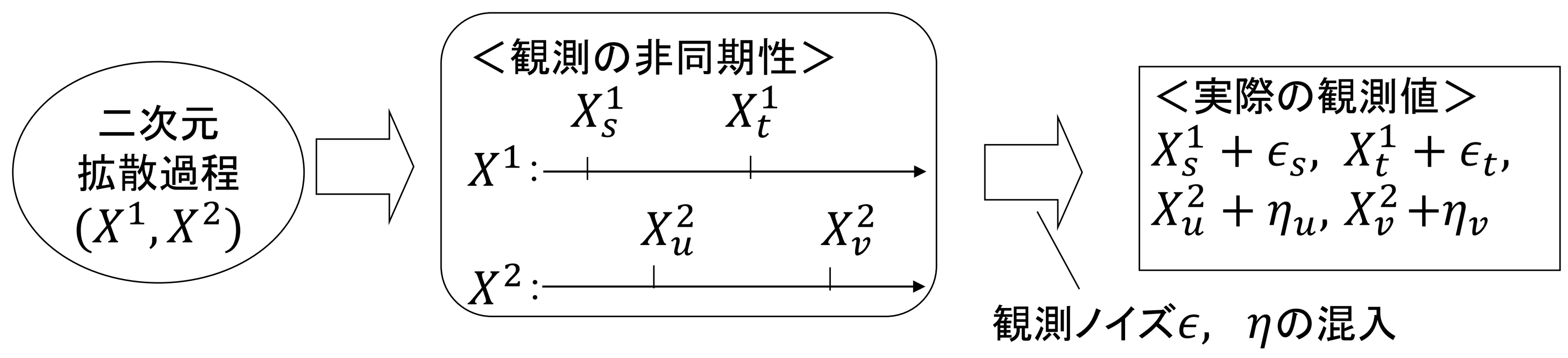
高頻度金融データは, そのデータの膨大さに加え, 特有の複雑な構造を持つため統計解析は容易ではない. ここでは拡散過程による株価モデリングを考える際の重要な問題として以下の二点を取り上げる.

i. **マーケット・マイクロストラクチャー・ノイズ**

　高頻度金融データにおける観測ノイズの混入が実証研究より示唆されている

ii. **非同期観測**

　異なる株式に対する観測時刻が異なる → 従来の統計解析手法の適用が困難




### 【拡散過程によるモデリング】

二次元の潜在株価過程 $X_t=(X_t^1, X_t^2)$ は確率微分方程式

$$dX_t=\mu(t,X_t,\sigma_*)dt+b(t,X_t,\sigma_*)dW_t, \qquad t\in[0,T]$$

を満たすとする. ただし, $W_t$は標準ブラウン運動, $\mu, b$は既知の関数, $\sigma_*$は$d$次元未知パラメータである. $X_t^1, X_t^2$ の観測時刻はそれぞれ$s_0^1, s_1^1, \dots, s_l^1$と$s_0^2, s_1^2, \dots, s_m^2$で与えられるとする（非同期観測）. さらに平均0の独立同分布の観測ノイズ$(\epsilon_i)_{0\le i\le l}$, $(\eta_j)_{0\le j\le m}$に対して, 観測データは

$$Y_i^1=X_{s_i^1}^1+\epsilon_i\ (0\le i\le l), \qquad Y_j^2=X_{s_j^2}^2+\eta_j\ (0\le j\le m)$$

で与えられるとする. 観測データから未知パラメータ$\sigma_*$を推定することで株価過程のボラティリティや共変動などのリスク量を計算することができる.

### 【最尤型推定法】

潜在株価過程$X_t$の局所ガウス近似を用いることにより尤度関数の近似$H_n(\sigma, v_1, v_2)$が得られ, 最尤型推定量$\hat{\sigma}_n$は$\hat{\sigma}_n=\mathrm{argmax}_\sigma H_n(\sigma, \hat{v}_{1,n}, \hat{v}_{2,n})$により得られる.

但し, $\hat{v}_{1,n}, \hat{v}_{2,n}$はそれぞれ$\epsilon_0, \eta_0$の分散$v_{1,*}, v_{2,*}$の推定量である.

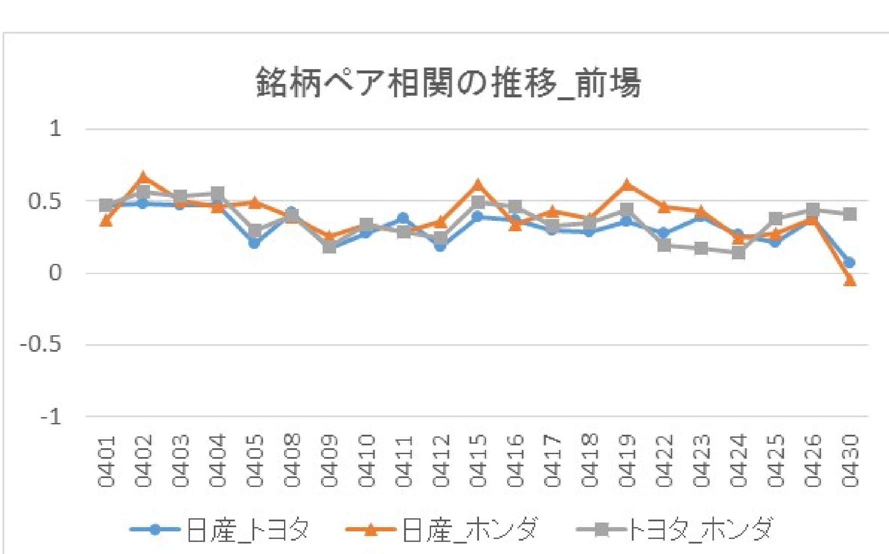
観測数のオーダーを$b_n$と書くと, 提案推定量$\hat{\sigma}_n$は一定の条件の下$n\to\infty$で以下の漸近混合正規性を満たす:

$$b_n^{1/4}(\hat{\sigma}_n-\sigma_*)\to^d \Gamma^{-1/2}\zeta.$$

ただし, $\Gamma$はある確率行列で, $\zeta$は$\Gamma$と独立な多変量標準正規乱数である.

### 【日本株式市場における分析】

日本株式市場の高頻度金融データ（日経NEEDS, 2013年4月）に対し, 上記のモデルを適用することで以下のような日内の株価相関の計算が可能である.



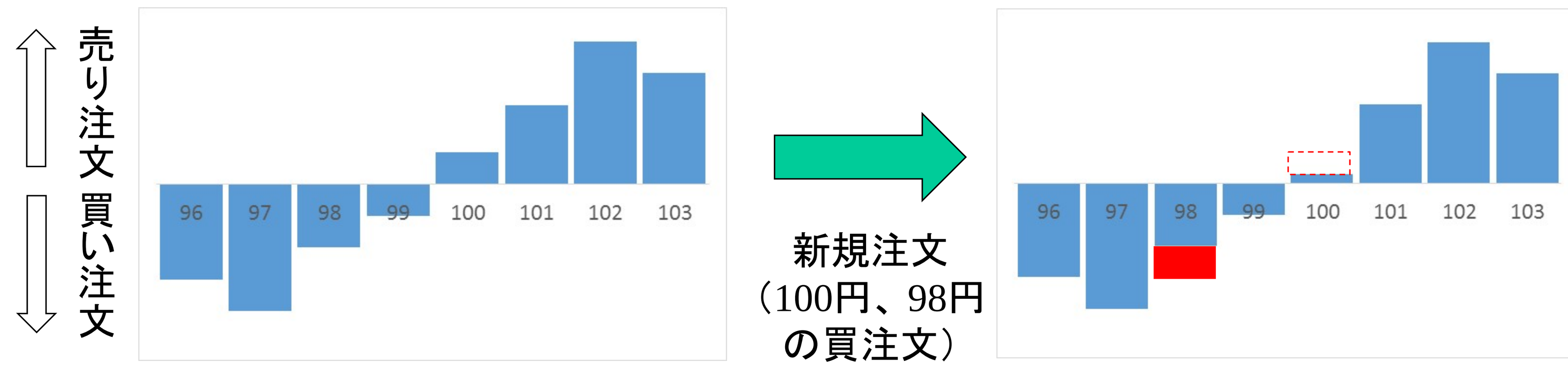

➢ 相関構造のダイナミクスを解析することが可能となり, 特に金融危機時等の相関構造の大きな変化に対応したリスク・コントロールが可能になる

## 3. 点過程を用いた株式板情報のモデリング

### 【株式板情報を用いた統計解析】

株式市場では, 各投資家が株を「いくらで何株買いたい（売りたい）」という注文をマッチングすることにより株が取引されている. 成立した取引のデータだけではなく, このような取引成立前の売買注文情報（株式板情報）を直接モデリングするという研究も行っている.

＜株式板情報＞




$N_t^p$を価格$p$における時刻tの注文総量とする. $N_t^p$は点過程と呼ばれる整数値の確率過程でモデリングされる.

### 【ホークス過程】

株式板情報のモデリングに用いられる点過程として、ホークス過程と呼ばれる自己励起型点過程の研究が進んでいる. 点過程$N_t$がホークス過程であるとは, その強度を$\lambda_t$と書いたときに,

$$\lambda_t=\gamma+\int_0^{t-}\alpha e^{-\beta(t-s)}dN_s, \qquad t\in(0,T)$$

を満たすことである. 但し, $\alpha, \beta, \gamma$は$\alpha<\beta$を満たす正定数とする.

$N_t$にジャンプが起きると, 強度$\lambda_t$は大きさ$\alpha$だけジャンプするので, よりジャンプが起きやすくなる（自己励起性）. ホークス過程のこのような性質により, 株式市場における注文のクラスタリング現象を表現することが可能になる.

### 【ホークス過程の最尤推定】

ホークス過程の対数尤度関数

$$l_n(\alpha,\beta,\gamma)=\int_0^T\log\lambda_t(\alpha,\beta,\gamma)\,dN_t-\int_0^T(\lambda_t(\alpha,\beta,\gamma)-1)dt$$

を最大にする$\alpha, \beta, \gamma$を求めることにより, 最尤推定量$\hat{\alpha}, \hat{\beta}, \hat{\gamma}$が計算される.

### 【漸近正規性】

自然数$n$に対し, $N_t^n$が強度$n\lambda_t^n$をもつとする. 但し,

$$\lambda_t^n=\gamma+\int_0^{t-}\alpha e^{-\beta(t-s)}n^{-1}dN_t^n, \qquad t\in(0,T).$$

この時パラメータ$\alpha, \beta, \gamma$のデータ$(N_t^n)_{0\le t\le T}$に対する最尤推定量を$\hat{\alpha}_n, \hat{\beta}_n, \hat{\gamma}_n$とおくと,

$$\sqrt{n}\left(\hat{\alpha}_n-\alpha, \hat{\beta}_n-\beta, \hat{\gamma}_n-\gamma\right)\to^d N\left(0, \tilde{\Gamma}^{-1/2}\right).$$

但し, $\tilde{\Gamma}$は$\alpha, \beta, \gamma$から定まる$3\times3$行列.

### 【応用例】

点過程を用いた株式板情報のモデリングの応用例として, 株式の大量執行時の株価変化の予測（マーケット・インパクト）がある. 年金基金等の機関投資家は大量の株式を保有しているため, 株式の売買執行が株価を大きく動かしてしまうという問題を抱えている. 株式板情報を点過程でモデリングすることで最適な執行戦略をシミュレーションから求めていくことが可能になると期待される.

株式板情報

| ask | price | bid |
|---|---|---|
| 583 | 10280 | |
| 582 | 10270 | |
| 642 | 10260 | |
| 757 | 10250 | |
| 685 | 10240 | |
| 769 | 10230 | |
| 742 | 10220 | |
| 829 | 10210 | |
| 887 | 10200 | |
| 1564 | 10190 | |
| | 10180 | 178 |
| | 10170 | 913 |
| | 10160 | 881 |
| | 10150 | 772 |
| | 10140 | 795 |
| | 10130 | 843 |
| | 10120 | 1017 |
| | 10110 | 679 |
| | 10100 | 649 |
| | 10090 | 580 |